Generalized Entropies

F. Dupuis L. Krämer P. Faist J. M. Renes R. Renner

Institute for Theoretical Physics
ETH Zurich, 8093 Zurich, Switzerland
{dupuis,lkraemer,pfaist,renes,renner}@phys.ethz.ch
www.qit.ethz.ch

13 November 2012

Abstract

We study an entropy measure for quantum systems that generalizes the *von Neumann entropy* as well as its classical counterpart, the *Gibbs* or *Shannon entropy*. After establishing a few basic properties of this generalized entropy, we show that it is closely related to *smooth entropies*, a family of entropy measures that is used to characterize a wide range of operational quantities.

1 Introduction

Entropy, originally introduced in thermodynamics, is nowadays recognized as a rather universal concept with a variety of uses, ranging from physics and chemistry to information theory and the theory of computation. Besides the role it plays for foundational questions, it is also relevant for applications. For example, entropy is used to study the efficiency of steam engines, but it also occurs in formulae for the data transmission capacity of optical fibres.

While entropy can be defined in various ways, a very common form employed for the study of classical systems is the *Gibbs entropy* or, in the context of information theory, the *Shannon entropy* [1]. It is defined for any probability distribution P as

$$H(P) = -\sum_x P(x) \log P(x)$$

(up to an unimportant proportionality factor). This definition has been generalized to the von Neumann entropy [2], which is defined for density operators,

$$H(\rho) = -\mathrm{Tr}(\rho \log \rho) \; .$$

While these entropy measures have a wide range of applications, it has recently become apparent that they are not suitable for correctly characterizing operationally relevant quantities in general scenarios (as explained below). This has led to the development of extensions [3], among them the *information spectrum approach* [4, 5, 6] and *smooth entropies* [7, 8] (where the former can be obtained as an asymptotic limit of the latter [9]).

The aim of this work is to study an alternative measure of entropy that generalizes von Neumann entropy. The generalized entropy is closely related to smooth entropies, which, in turn, are connected to a variety of operational quantities.

1.1 Axiomatic and operational approach to entropy

The variety of areas and applications where entropies are used is impressive, and one may wonder what it is that makes entropy such a versatile concept.

One may attempt to answer the question from an *axiomatic* viewpoint. Here, the idea is to consider (small) sets of axioms that characterize the nature of entropy. There is a vast amount of literature devoted to the specification of such axioms and their study [1, 10, 11, 12, 13, 14, 15, 16]. While the choice of a set of axioms is ultimately a matter of taste, we sketch in the following some of the most popular axioms. We do this for the case of entropies defined on quantum systems, i.e., we consider functions H from the set of density operators (denoted by ρ) to the real numbers.

- *Positivity:* $H(\rho) \geqslant 0$.
- *Invariance under isometries:* $H(U\rho U^\dagger) = H(\rho)$.
- *Continuity:* H is a continuous function of ρ.
- *Additivity:* $H(\rho_A \otimes \rho_B) = H(\rho_A) + H(\rho_B)$.
- *Subadditivity:* $H(\rho_{AB}) \leqslant H(\rho_A) + H(\rho_B)$.[1]

The (special) case of classical entropies is obtained by replacing the density operators by probability distributions. Note that the second axiom then reduces to the requirement that the entropy is invariant under permutations.

It is easy to verify that the von Neumann entropy satisfies the above axioms. Furthermore, it can be shown that (up to a constant factor, which may be fixed by an additional normalization axiom), the von Neumann entropy is essentially the only function satisfying the above postulates [12]. This result—as well as similar results based on slightly different sets of axioms—nicely expose the universal nature of entropy. Note, in particular, that the above axioms do not refer specifically to thermodynamic or information-theoretic properties of a system.

An alternative to this axiomatic approach is to relate entropy to *operational* quantities. In thermodynamics, examples for such operational quantities include measures for heat flow or the amount of work that is transformed into heat during a given process. In information theory, operational quantities are, for instance, the minimum size to which the information generated by a source can be compressed, or the amount of uniform randomness that can be extracted from a non-uniform source.

Given the very different nature of these operational quantities, it is not obvious that this approach can lead to a reasonable notion of entropy. One would rather expect an entire family of entropy measures—possibly as large as the number of different operational quantities one considers. However, there exist remarkable connections, even relating thermodynamic and information-theoretic quantities. For example, it follows from Landauer's principle [17, 18] that the amount of work that can be extracted from a system is directly related to the size to which the information contained in it can be compressed [19, 20, 21].

[1] Here ρ_{AB} denotes a density operator on a bipartite system and ρ_A and ρ_B are obtained by partial traces over the second and first subsystem, respectively.

Recent work has shown that a large number of operational quantities can be characterized with one single class of entropy measures. *Smooth entropies* (denoted by $H_{\min}^{\epsilon}$ and $H_{\max}^{\epsilon}$), which were developed mostly within quantum information theory, are an example of such a class. For instance, $H_{\min}^{\epsilon}$ quantifies the number of uniformly random (classical) bits that can be deterministically extracted from a weak source of randomness [8, 22] and $H_{\max}^{\epsilon}$ quantifies the number of bits needed to encode a given (classical) value [23]. More generally, $H_{\min}^{\epsilon}$ can be used to characterize *decoupling* [24], a quantum version of randomness extraction [25], and *state merging* [26, 27], which can be seen as the fully quantum analogue of coding [28]. Also, a combination of $H_{\min}^{\epsilon}$ and $H_{\max}^{\epsilon}$ gives an expression for the classical capacity of a classical [29] or a quantum [30] channel, as well as its "reverse" capacity [31]. Additional applications can be found particularly in quantum cryptography (see, e.g., [8, 32, 33]). Smooth entropies also have operational interpretations within thermodynamics. For example, they can be used in a single-shot version of Landauer's principle to quantify the amount of work required by an operation that moves a given system into a pure state [19, 20, 21].

However, smooth entropies are generally different from the von Neumann entropy except in special cases. This implies that many operational quantities, characterized by smooth entropies, are not in general accurately described by the von Neumann entropy (e.g. the amount of extractable randomness or the encoding length). In particular, it follows that some of the axioms considered above must be incompatible with the operational approach.

This can also be seen directly, for example, for the (classical) task of randomness extraction. Let $C(X)$ be the number of uniform bits that can be obtained by applying a function to a random variable X distributed according to P_X. Then the quantity C automatically has the properties one would expect from an uncertainty measure: it equals 0 if X is perfectly known, and it increases as X becomes more uncertain. One may therefore interpret C as an (operationally defined) entropy measure for classical random variables.

However, while C is indeed positive, invariant under permutations, and additive, it is not subadditive. To see this, consider a random variable R uniformly distributed over the set $\{1, \ldots, 2^{\ell}\}$, for some large $\ell \in \mathbb{N}$. Furthermore, define the random variables X and Y by

$$X = \begin{cases} R & \text{if } R \leqslant 2^{\ell-1} \\ 0 & \text{otherwise} \end{cases}$$
$$Y = \begin{cases} R & \text{if } R > 2^{\ell-1} \\ 0 & \text{otherwise.} \end{cases}$$

Since $\Pr[X = 0] = \Pr[Y = 0] = \frac{1}{2}$, it is not possible to extract more than 1 bit from either of X or Y separately, i.e., $C(X) = C(Y) \leqslant 1$. However, since the pair (X, Y) is in one-to-one relation to R, we have $C(XY) = C(R) = \ell$. Hence, subadditivity, $C(XY) \leqslant C(X) + C(Y)$ can be violated by an arbitrarily large amount. [2]

1.2 Generalized entropy measure

The above considerations show that an operational approach to entropies necessitates the use of entropy measures that are more general than those obtained by the usual axiomatic

[2] However, an inequality of similar form can be recovered — this is known as *e*ntropy splitting lemma [34, 35]

approaches. The aim of this paper is to investigate such a generalization. The entropy measure we consider is motivated by previous work [36, 37, 38, 39].

We consider a family of entropies, denoted H_H^ϵ, parametrized by a real number ϵ from the interval $[0, 1]$. H_H^ϵ is defined via a relative-entropy type quantity, i.e., a function that depends on two density operators, ρ and σ, similarly to the Kullback-Leibler divergence [40, 41]. This quantity, denoted D_H^ϵ, has a simple interpretation in the context of quantum hypothesis testing [42]. Consider a measurement for distinguishing whether a system is in state ρ or σ. $D_H^\epsilon(\rho\|\sigma)$ then corresponds to the negative logarithm of the failure probability when the system is in state σ, under the constraint that the success probability when the system is in state ρ is at least ϵ (see Section 3.1 below).

Starting from $D_H^\epsilon(\rho\|\sigma)$, it is possible to directly define a *conditional entropy*, $H_H^\epsilon(A|B)$, i.e., a measure for the uncertainty of a system A conditioned on a system B (see Section 3.2 below). We note that, while the conditional von Neumann entropy may be defined analogously using the Kullback-Leibler divergence, the standard expression for conditional von Neumann entropy [43],

$$H(A|B) = H(\rho_{AB}) - H(\rho_B) \,, \tag{1}$$

cannot be generalized directly. However, as shown in Section 5, H_H^ϵ satisfies a *chain rule*, i.e., an inequality which resembles (1). In addition, we show that H_H^ϵ has many desirable properties that one would expect an entropy measure to have (see Section 3.3), for instance that it reduces to the von Neumann entropy in the asymptotic limit (Asymptotic Equipartition Property).

A central part of this contribution is to establish direct relations to the *smooth entropy measures* $H_{\min}^\epsilon$ and $H_{\max}^\epsilon$ (Section 4). As explained above, it has been shown that these accurately characterize a number of operational quantities, such as information compression, randomness extraction, entanglement manipulation, and channel coding. Furthermore, they are also relevant in the context of thermodynamics, e.g., for quantifying the amount of work that can be extracted from a given system. The bounds derived in Section 4 imply that H_H^ϵ has a similar operational significance.

2 Preliminaries

2.1 Notation and Definitions

For a finite-dimensional Hilbert space $\mathcal{H}$, let $\mathcal{L}(\mathcal{H})$ and $\mathcal{P}(\mathcal{H})$ be the linear and positive semi-definite operators on $\mathcal{H}$, respectively. On $\mathcal{L}(\mathcal{H})$ we employ the Hilbert-Schmidt inner product $\langle X, Y\rangle := \mathrm{Tr}(X^\dagger Y)$. Quantum states form the set $\mathcal{S}(\mathcal{H}) = \{\rho \in \mathcal{P}(\mathcal{H}) : \mathrm{Tr}(\rho) = 1\}$, and we define the set of sub-normalized states as $\mathcal{S}_\leqslant(\mathcal{H}) = \{\rho \in \mathcal{P}(\mathcal{H}) : 0 < \mathrm{Tr}(\rho) \leqslant 1\}$. To describe multi-partite quantum systems on tensor product spaces we use capital letters and subscripts to refer to individual subsystems or marginals. We call a state ρ_{XB} *classical-quantum (CQ)* if it is of the form $\rho_{XB} = \sum_x p(x) \left|x\right\rangle\left\langle x\right| \otimes \rho_B^x$ with $\rho_B^x \in \mathcal{S}(\mathcal{H}_B)$, $p(x)$ a probability distribution and $\{|x\rangle\}$ an orthonormal basis of $\mathcal{H}_X$.

A map $\mathcal{E} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$ which, for any $\mathcal{H}''$, maps $\mathcal{P}(\mathcal{H} \otimes \mathcal{H}'')$ to $\mathcal{P}(\mathcal{H}' \otimes \mathcal{H}'')$ is called a completely positive map (CPM). It is called trace-preserving if $\mathrm{Tr}(\mathcal{E}[X]) = \mathrm{Tr}(X)$ for any $X \in \mathcal{P}(\mathcal{H})$. A unital map satisfies $\mathcal{E}(\mathbb{I}) = \mathbb{I}$, and a map is sub-unital if $\mathcal{E}(\mathbb{I}) \leqslant \mathbb{I}$. The adjoint $\mathcal{E}^*$ of $\mathcal{E}$ is defined by $\mathrm{Tr}\left(\mathcal{E}^*(Y)\, X\right) = \mathrm{Tr}\left(Y\, \mathcal{E}(X)\right)$.

We employ two distance measures on sub-normalized states: the purified distance $P(\rho, \sigma)$ [44, 45, 46] and the trace distance $D(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1$ (where $||\rho||_1 = \mathrm{Tr}(\sqrt{\rho^\dagger \rho})$). The purified distance is defined in terms of the fidelity $F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1$ by $P(\rho, \sigma) = \sqrt{1 - F(\rho, \sigma)^2}$. The purified and trace distances obey the following relation [47]: $D(\rho, \sigma) \leqslant P(\rho, \sigma) \leqslant \sqrt{2D(\rho, \sigma)}$.

Finally, the operator inequality $A \leqslant B$ is taken to mean that $A - B$ is positive semi-definite, and when comparing a matrix to a scalar we assume that the scalar is multiplied by the identity matrix. Note also that all logarithms taken in the calculations are base 2.

2.2 Semi-Definite Programs

Watrous has given an elegant formulation of semidefinite programs especially adapted to the present context [48]. Here we follow his notation; see also [49] for a more extensive treatment. A semidefinite program over $\mathcal{X} = \mathbb{C}^n$ and $\mathcal{Y} \in \mathbb{C}^m$ is specified by a triple (Ψ, A, B), for A and B Hermitian operators in $\mathcal{L}(\mathcal{X})$ and $\mathcal{L}(\mathcal{Y})$ respectively, and $\Psi : \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{Y})$ a linear, Hermiticity-preserving operation.

This semidefinite program corresponds to two optimization problems, the so-called "primal" and "dual" problems:

PRIMAL		DUAL	
minimize	$\langle A, X \rangle$	maximize	$\langle B, Y \rangle$
subj. to	$\Psi(X) \geqslant B$	subj. to	$\Psi^*(Y) \leqslant A$
	$X \in \mathcal{P}(\mathcal{X})$		$Y \in \mathcal{P}(\mathcal{Y})$

With respect to these problems, one can define the primal and dual feasible sets $\mathcal{A}$ and $\mathcal{B}$ respectively:

$$\mathcal{A} = \{X \in \mathcal{P}(\mathcal{X}) : \Psi(X) \geqslant B\}, \tag{2}$$

$$\mathcal{B} = \{Y \in \mathcal{P}(\mathcal{Y}) : \Psi^*(Y) \leqslant A\}. \tag{3}$$

The operators $X \in \mathcal{A}$ and $Y \in \mathcal{B}$ are then called primal and dual feasible (solutions) respectively.

To each of the primal and dual problems, the associated optimal values are defined as:[3]

$$\alpha = \inf_{X \in \mathcal{A}} \langle A, X \rangle \quad \text{and} \quad \beta = \sup_{Y \in \mathcal{B}} \langle B, Y \rangle .$$

Solutions to the primal and dual problems are related by the following two duality theorems:

Theorem 1. *(Weak duality). $\alpha \geqslant \beta$ for every semidefinite program (Ψ, A, B).*

Theorem 2. *(Slater-type condition for strong duality). For every semi-definite program (Ψ, A, B) as defined above, the following two statements hold:*

1. *Strict primal feasibility: If β is finite and there exists an operator $X > 0$ s.t. $\Psi(X) > B$, then $\alpha = \beta$ and there exists $Y \in \mathcal{B}$ s.t. $\langle B, Y \rangle = \beta$.*

2. *Strict dual feasibility: If α is finite and there exists an operator $Y > 0$ s.t. $\Psi^*(Y) < A$, then $\alpha = \beta$ and there exists $X \in \mathcal{A}$ s.t. $\langle A, X \rangle = \alpha$.*

[3] If $\mathcal{A} = \varnothing$ or $\mathcal{B} = \varnothing$, we define $\alpha = \infty$ or $\beta = -\infty$ respectively

Given strict feasibility, we obtain *complementary slackness* conditions linking the optimal X and Y for primal and dual problems:

$$\Psi(X)Y = BY \quad \text{and} \quad \Psi^*(Y)X = AX. \tag{4}$$

Semidefinite programs can be solved efficiently using the ellipsoid method [50]. There exists an algorithm that, under certain stability conditions and bounds on the primal feasible and dual feasible sets, finds an approximation for the optimal value of the primal problem. The running time of the algorithm is bounded by a polynomial in n, m, and the logarithm of the desired accuracy (see [48] for more details).

3 Relative and Conditional Entropies

We will now introduce the new family of entropy measures, as well as the smooth entropies, and the set of relative entropies that they are based on.

3.1 Definition of relative entropies

We define the ϵ-relative entropy $D_H^\epsilon(\rho||\sigma)$ of a normalized state $\rho \in \mathcal{S}(\mathcal{H})$ relative to $\sigma \in \mathcal{P}(\mathcal{H})$ as [4]

$$2^{-D_H^\epsilon(\rho||\sigma)} := \frac{1}{\epsilon} \inf\{\langle Q, \sigma\rangle \,|0 \leqslant Q \leqslant 1 \wedge \langle Q, \rho\rangle \geqslant \epsilon\}\,. \tag{5}$$

This corresponds to minimizing the probability that a strategy Q to distinguish ρ from σ produces a wrong guess on input σ while maintaining a minimum success probability ϵ to correctly identify ρ. In particular, for $\epsilon = 1$, $D_H^\epsilon(\rho||\sigma)$ is equal to Rényi's entropy [51] of order 0, and $D_0(\rho||\sigma) = -\log \mathrm{Tr}(\rho^0\sigma)$, with ρ^0 the projector on the support of ρ [39].

The relative min- and max-entropies $D_{\min}$ and $D_{\max}$ for $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ are defined as follows:

$$2^{-D_{\min}(\rho||\sigma)} = F(\rho, \sigma)^2, \tag{6}$$

$$D_{\max}(\rho||\sigma) = \inf\{\lambda \in \mathbb{R} : 2^\lambda \sigma \geqslant \rho\}. \tag{7}$$

These definitions can easily be extended to subnormalized states $\rho, \sigma \in \mathcal{S}_\leqslant(\mathcal{H})$ by using the generalized fidelity $F(\rho, \sigma) = \mathrm{Tr}|\sqrt{\rho}\sqrt{\sigma}| + \sqrt{(1 - \mathrm{Tr}\rho)(1 - \mathrm{Tr}\sigma)}$. We define also the corresponding smoothed quantities:

$$D_{\min}^\epsilon(\rho||\sigma) = \max_{\tilde{\rho} \in \mathcal{B}_\epsilon(\rho)} D_{\min}(\tilde{\rho}||\sigma), \tag{8}$$

$$D_{\max}^\epsilon(\rho||\sigma) = \min_{\tilde{\rho} \in \mathcal{B}_\epsilon(\rho)} D_{\max}(\tilde{\rho}||\sigma), \tag{9}$$

with $\mathcal{B}_\epsilon(\rho) = \{\tilde{\rho} \in \mathcal{S}_\leqslant(\mathcal{H}) | P(\tilde{\rho}, \rho) \leqslant \epsilon\}$ the purified-distance-ball around ρ so that the optimization is over all subnormalized states $\tilde{\rho}$ ϵ-close to ρ with respect to the purified distance.

[4] Note that this differs slightly from both the definitions used by Wang and Renner [38] and by Tomamichel and Hayashi [39]. Similar formulations specific to mutual information and entanglement were previously given respectively by Buscemi and Datta [36] and Brandão and Datta [37].

3.2 Definition of the conditional entropies

We define the new entropy $H_H^\epsilon(A|B)_\rho$, in terms of the relative entropy we have already introduced, as follows:

$$H_H^\epsilon(A|B)_\rho := -D_H^\epsilon(\rho_{AB}||\mathbb{I}_A \otimes \rho_B) \tag{10}$$

In the smooth entropy framework, the min- and max- entropies are given by: [46, 52, 53]

$$H_{\min}^\epsilon(A|B)_{\rho|\sigma} := -D_{\max}^\epsilon(\rho_{AB}\|\mathbb{I}_A \otimes \sigma_B)\,, \tag{11}$$

$$H_{\max}^\epsilon(A|B)_{\rho|\sigma} := -D_{\min}^\epsilon(\rho_{AB}\|\mathbb{I}_A \otimes \sigma_B)\,, \tag{12}$$

$$H_{\min}^\epsilon(A|B)_\rho := \max_{\tilde\rho\in\mathcal{B}_\epsilon(\rho)} \sup_{\sigma_B} -D_{\max}(\tilde\rho_{AB}\|\mathbb{I}_A \otimes \sigma_B)\,, \tag{13}$$

$$H_{\max}^\epsilon(A|B)_\rho := \min_{\tilde\rho\in\mathcal{B}_\epsilon(\rho)} \sup_{\sigma_B} -D_{\min}(\tilde\rho_{AB}\|\mathbb{I}_A \otimes \sigma_B)\,. \tag{14}$$

The non-smoothed versions $H_{\min}(A|B)$ and $H_{\max}(A|B)$ are given by taking $\epsilon = 0$.

For the special case when $\epsilon \to 0$, $H_H^\epsilon(A|B)$ converges to $H_{\min}(A|B)_{\rho|\rho}$ since for the optimal solutions to the semi-definite program as defined below $X \to 0$. In the case where one is also not conditioning on any B-system (i.e. take B to be a trivial system, or take $\rho_{AB} = \rho_A \otimes \rho_B$), then H_H^ϵ reduces to the min-entropy:

$$\lim_{\epsilon\to 0} H_H^\epsilon(A)_\rho = H_{\min}(A)_\rho = -\log||\rho_A||_\infty. \tag{15}$$

Note also that H_H^ϵ is monotonically increasing in ϵ: to see this, find that the dual optimal $\{\mu, X\}$ for $2^{H_H^\epsilon}$ (see below) is also feasible for $2^{H_H^{\epsilon'}}$ with $\epsilon' \geqslant \epsilon$.

3.3 Elementary Properties

As we are going to show in this section, the quantities D_H^ϵ and H_H^ϵ we introduced satisfy many desirable properties one would expect from an entropy measure.

3.3.1 Properties of D_H^ϵ

D_H^ϵ can be expressed in terms of a semi-definite program, meaning it can be efficiently approximated. Due to strong duality we obtain two equivalent expressions with optimal solutions linked by complementary slackness conditions [49]. The semi-definite program for $2^{-D_H^\epsilon(\rho||\sigma)}$ reads:

PRIMAL		DUAL	
minimize	$\frac{1}{\epsilon}$Tr[Qσ]	maximize	$\mu - \frac{\mathrm{Tr}[X]}{\epsilon}$
subj. to	Q$\leqslant \mathbb{I}$	subj. to	$\mu\rho \leqslant \sigma$+X
	Tr[Qρ]$\geqslant \epsilon$		$X \geqslant 0$
	$Q \geqslant 0$		

This yields the following complementary slackness conditions for primal and dual optimal solutions $\{Q\}$ and $\{\mu, X\}$:

$$(\mu\rho - X)Q = \sigma Q \tag{16}$$
$$\mathrm{Tr}[Q\rho] = \epsilon \tag{17}$$
$$QX = X \tag{18}$$

from which we can infer that $[Q, X] = 0$, as well as the fact that the positive part of $(\mu\rho - \sigma)$ is in the eigenspace of Q with eigenvalue 1.

Further properties include:

Proposition 1 (Positivity)**.** *For any $\rho, \sigma \in \mathcal{S}(\mathcal{H})$,*

$$D_H^\epsilon(\rho||\sigma) \geqslant 0, \tag{19}$$

with equality if $\rho = \sigma$.

Proof. Positivity follows immediately from the definition of D_H^ϵ by choosing $Q = \epsilon\mathbb{I}$. Equality is achieved if $\rho = \sigma$ because $\frac{1}{\epsilon}\min_{\mathrm{Tr}(Q\rho)\geqslant\epsilon}\mathrm{Tr}(Q\rho) = 1$. □

Note that $D_H^\epsilon(\rho\|\sigma) = 0$ does not generally imply $\rho = \sigma$: for example, consider the case where $\epsilon = 1$ and where ρ and σ have same support.

Proposition 2 (Data Processing Inequality (DPI))**.** *For any completely positive, trace non-increasing map $\mathcal{E}$,*

$$D_H^\epsilon(\rho||\sigma) \geqslant D_H^\epsilon(\mathcal{E}(\rho)||\mathcal{E}(\sigma)). \tag{20}$$

Proof. For a proof of this DPI, see [38]. □

Proposition 3 (Asymptotic Equipartition Property)**.** *Let*

$$D(\rho||\sigma) = \mathrm{Tr}[\rho(\log\rho - \log\sigma)]$$

be the relative entropy between ρ and σ [41]. Then, for any $0 < \epsilon \leqslant 1$,

$$\lim_{n\to\infty}\frac{1}{n}D_H^\epsilon(\rho^{\otimes n}||\sigma^{\otimes n}) = D(\rho||\sigma). \tag{21}$$

Proof. From Stein's lemma [3, 54] it immediately follows that

$$\lim_{n\to\infty}\frac{1}{n}D_H^\epsilon(\rho^{\otimes n}||\sigma^{\otimes n}) = \lim_{n\to\infty} -\frac{1}{n}\log\min\frac{1}{\epsilon}\mathrm{Tr}\{\sigma^{\otimes n}Q\}, \tag{22}$$
$$= D(\rho||\sigma) - \lim_{n\to\infty}\frac{1}{n}(\log\frac{1}{\epsilon}) \tag{23}$$
$$= D(\rho||\sigma), \tag{24}$$

where the minimum is taken over $0 \leqslant Q \leqslant 1$ such that $\mathrm{Tr}Q\rho \geqslant \epsilon$. □

3.3.2 Properties of H_H^ϵ

Proposition 4 (Bounds)**.** *For ρ_{AB} an arbitrary normalized quantum state and ρ_{XB} a classical-quantum state,*

$$-\log|A| \leqslant H_H^\epsilon(A|B)_\rho \leqslant \log|A|, \tag{25}$$

$$0 \leqslant H_H^\epsilon(X|B)_\rho \leqslant \log|X|. \tag{26}$$

For classical-quantum states, $H_H^\epsilon(X|B) = 0$ if X is completely determined by B (so that $\mathrm{Tr}(\rho_B^x \rho_B^{x'}) = 0$ for any $x' \neq x$), and the entropy is maximal if X is completely mixed and independent of B (i.e. $\rho_{XB} = \frac{1}{|X|}\mathbb{I}_X \otimes \rho_B$).

Proof. Start with the upper bound on H_H^ϵ, and choose $\epsilon\mathbb{I}$ as a feasible Q:

$$2^{H_H^\epsilon(A|B)_\rho} = \min_{\mathrm{Tr}[Q_{AB}\rho_{AB}]\geqslant\epsilon} \frac{1}{\epsilon}\mathrm{Tr}[Q_{AB}\mathbb{I}_A \otimes \rho_B] \tag{27}$$

$$\leqslant \frac{1}{\epsilon}\mathrm{Tr}[\epsilon\mathbb{I}_{AB}\mathbb{I}_A \otimes \rho_B] \tag{28}$$

$$= |A|. \tag{29}$$

For the lower bound we use the inequality $|A|\mathbb{I}_A \otimes \rho_B \geqslant \rho_{AB}$, which holds for arbitrary quantum states ρ_{AB}. To establish this inequality, define the superoperator $\mathcal{E}$ as $\mathcal{E}(\rho) = \frac{1}{d^2}\sum_{j,k}(U^jV^k)\rho(U^jV^k)^\dagger$. Here, $d = \dim(\mathcal{H})$ while U and V are unitary operators defined by $U|j\rangle = |j+1\rangle$ and $V|k\rangle = \omega^k|k\rangle$, for an orthonormal basis set $\{|j\rangle\}_{j=0}^{d-1}$, $\omega = e^{2\pi i/d}$, and where arithmetic inside the ket is taken modulo d. (The operators U and V are often called the discrete Weyl-Heisenberg operators, as they generate a discrete projective representation of the Heisenberg algebra.) Then it is easy to work out that $\mathcal{E}\otimes\mathbb{I}[\rho^{AB}] = \frac{1}{|A|}\mathbb{I}_A\otimes\rho_B$, which by the form of $\mathcal{E}$ implies the sought-after inequality. Then, for the optimal Q_{AB} in $H_H^\epsilon(A|B)_\rho$,

$$2^{H_H^\epsilon(A|B)_\rho} = \frac{1}{\epsilon}\mathrm{Tr}[Q_{AB}\mathbb{I}_A \otimes \rho_B] \tag{30}$$

$$\geqslant \frac{1}{\epsilon|A|}\mathrm{Tr}[Q_{AB}\rho_{AB}] \tag{31}$$

$$\geqslant \frac{1}{|A|}. \tag{32}$$

Classical-quantum states ρ_{XB} obey $\mathbb{I}_X \otimes \rho_B \geqslant \rho_{XB}$, as $\sum_{x'} p_{x'}\rho_B^{x'} \geqslant p_x\rho_B^x$ for all x. This implies $H_H^\epsilon(X|B)_\rho \geqslant 0$ by the same argument.

That the extremal cases are reached for the described cases follows immediately from the respective definitions of ρ_{XB} and H_H^ϵ. □

Similarly to D_H^ϵ, H_H^ϵ also satisfies a data processing inequality[5].

Proposition 5 (Data Processing Inequality)**.** *For any $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{\mathcal{AB}})$, let $\mathcal{E} : A \to A'$ be a sub-unital TP-CPM, and $\mathcal{F} : B \to B'$ be a TP-CPM. Then, for $\tau_{A'B'} = \mathcal{E}\circ\mathcal{F}(\rho_{AB})$,*

$$H_H^\epsilon(A|B)_\rho \leqslant H_H^\epsilon(A'|B')_\tau \tag{33}$$

[5]This proof is adapted from the DPI proof for a differently defined H^ϵ in Tomamichel and Hayashi [39]

Proof. Let $\{\mu, X_{AB}\}$ be dual-optimal for $H_H^\epsilon(A|B)_\rho$. Starting from $\mu\rho_{AB} \leqslant \mathbb{I}_A \otimes \rho_B + X_{AB}$ and applying $\mathcal{E} \circ \mathcal{F}$ to both sides of the inequality yields:

$$\mu\tau_{AB} \leqslant \mathcal{E}(\mathbb{I}_A) \otimes \tau_{B'} + \mathcal{E} \circ \mathcal{F}(X_{AB}) \leqslant \mathbb{I}_{A'} \otimes \tau_{B'} + \mathcal{E} \circ \mathcal{F}(X_{AB}). \tag{34}$$

Hence, $\{\mu, \mathcal{E}\circ\mathcal{F}(X_{AB})\}$ is dual feasible for $H_H^\epsilon(A'|B')_\tau$ and $2^{H_H^\epsilon(A'|B')_\tau} \geqslant \mu - \mathrm{Tr}(\mathcal{E}\circ\mathcal{F}(X_{AB})/\epsilon) = 2^{H_H^\epsilon(A|B)_\rho}$. □

Proposition 6 (Asymptotic Equipartition Property)**.** *For any $0 < \epsilon \leqslant 1$, it holds that*

$$\lim_{n\to\infty} \frac{1}{n} H_H^\epsilon(A^n|B^n)_{\rho^{\otimes n}} = H(A|B)_\rho \,, \tag{35}$$

where $H(A|B)$ refers to the conditional von Neumann entropy.

Proof. Using the asyptotic property of D_H^ϵ derived from Stein's lemma above, we can show for $H_H^\epsilon(A|B)$:

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}(H_H^\epsilon(A^{\otimes n}|B^{\otimes n})_\rho) &= \lim_{n\to\infty} \frac{1}{n}(-D_H^\epsilon(\rho^{\otimes n}||(\mathbb{I}_A \otimes \rho_B)^{\otimes n})) && (36)\\
&= -D(\rho_{AB}||\mathbb{I}_A \otimes \rho_B) && (37)\\
&= -\mathrm{Tr}\rho_{AB}(\log \rho_{AB} - \log \mathbb{I}_A \otimes \rho_B) && (38)\\
&= H(AB) - \mathrm{Tr}(\rho_B \log \rho_B) && (39)\\
&= H(AB) - H(B) && (40)\\
&= H(A|B) \,. && (41)
\end{aligned}$$

□

4 Relation to (relative) min- and max-entropies

The following propositions relate the new quantities to smooth entropies. This guarantees an operational significance for D_H^ϵ and H_H^ϵ (see Section 1.1).

Proposition 7. *Let $\rho \in \mathcal{S}(\mathcal{H}_{\mathcal{AB}})$, $\sigma \in \mathcal{P}(\mathcal{H}_{\mathcal{AB}})$ and $0 < \epsilon \leqslant 1$. Then,*

$$D_{max}^{\sqrt{2\epsilon}}(\rho||\sigma) \leqslant D_H^\epsilon(\rho||\sigma) \leqslant D_{max}(\rho||\sigma) \tag{42}$$

$$H_{min}^{\sqrt{2\epsilon}}(A|B)_\rho \geqslant H_H^\epsilon(A|B)_\rho \geqslant H_{min}(A|B)_{\rho|\rho} \tag{43}$$

Proof. The upper bound for D_H^ϵ follows immediately from the fact that $\mu = 2^{-D_{\max}(\rho||\sigma)}$ and $X = 0$ are feasible for $2^{-D_H^\epsilon(\rho||\sigma)}$ in the dual formulation. For the lower bound, let μ and X be dual-optimal for $2^{-D_H^\epsilon(\rho||\sigma)}$. Now define $G := \sigma^{1/2}(\sigma + X)^{-1/2}$ and let $\tilde{\rho} := G\rho G^\dagger$. It thus follows that $\mu\tilde{\rho} \leqslant \sigma$, and hence $2^{-D_{\max}(\tilde{\rho}||\sigma)} \geqslant \mu$. Since $\mathrm{Tr}[X] \geqslant 0$, it holds that $\mu \geqslant 2^{-D_H^\epsilon(\rho||\sigma)}$, which implies that $2^{-D_H^\epsilon(\rho||\sigma)} \leqslant 2^{-D_{\max}(\tilde{\rho}||\sigma)}$.

It is now left to prove that the purified distance between $\tilde{\rho}$ and ρ does not exceed $\sqrt{2\epsilon}$: For this we employ Lemma 3, from which we obtain the upper bound $\sqrt{\frac{2}{\mu}\mathrm{Tr}[X]}$. Together with $0 \leqslant \epsilon\mu - \mathrm{Tr}[X]$, this implies that $P(\rho, \tilde{\rho}) \leqslant \sqrt{2\epsilon}$, which concludes the proof.

These bounds can now be rewritten to relate H_H^ϵ to $H_{\min}^\epsilon$. We have

$$H_{\min}^{\sqrt{2\epsilon}}(A|B)_\rho \geqslant -D_{\max}^{\sqrt{2\epsilon}}(\rho_{AB}||\mathbb{I}_A \otimes \rho_B) \geqslant -D_H^\epsilon(\rho_{AB}||\mathbb{I}_A \otimes \rho_B) = H_H^\epsilon(A|B)_\rho. \tag{44}$$

In the other direction we find:

$$H_H^\epsilon(A|B)_\rho = -D_H^\epsilon(\rho_{AB}||\mathbb{I}_A \otimes \rho_B) \geqslant -D_{\max}(\rho_{AB}||\mathbb{I}_A \otimes \rho_B) := H_{\min}(A|B)_{\rho|\rho}. \tag{45}$$

□

Proposition 8. *Let $\rho \in \mathcal{S}(\mathcal{H})$ and $\sigma \in \mathcal{P}(\mathcal{H})$ have intersecting support, and $0 < \epsilon \leqslant 1$. Then,*

$$D_{min}(\rho||\sigma) - \log\frac{1}{\epsilon^2} \leqslant D_H^{1-\epsilon}(\rho||\sigma) \leqslant D_{min}^{\sqrt{2\epsilon}}(\rho||\sigma) - \log\frac{1}{(1-\epsilon)} \tag{46}$$

$$H_{max}(A|B)_\rho + \log\frac{1}{\epsilon^2} \geqslant H_H^{1-\epsilon}(A|B)_\rho \tag{47}$$

Proof. We begin with the lower bound for $D_H^{1-\epsilon}$. Let μ, Q, and X be optimal for the primal and dual programs for $2^{-D_H^{1-\epsilon}(\rho||\sigma)}$ and define $Q^\perp := 1 - Q$. Complementary slackness implies $\mathrm{Tr}[Q^\perp\rho] = \epsilon$, $QX = X$ and $Q(\mu\rho - \sigma - X) = 0$. Thus,

$$Q(\mu\rho - \sigma - X) = Q(\mu\rho - \sigma) - X, \tag{48}$$

meaning $Q(\mu\rho - \sigma)$ is hermitian and positive semidefinite. This implies that $Q^\perp(\mu\rho - \sigma)$ is also hermitian and $Q^\perp(\mu\rho - \sigma) \leqslant 0$. Since $Q + Q^\perp = \mathbb{I}$, this gives a decomposition of $(\mu\rho - \sigma)$ into positive and negative parts, and thus $|\mu\rho - \sigma| = Q(\mu\rho - \sigma) - Q^\perp(\mu\rho - \sigma)$. We can now proceed:

$$\begin{aligned}
2^{-\frac{1}{2}D_{\min}(\rho||\sigma)} &= F(\rho, \sigma) && (49)\\
&= \frac{1}{\sqrt{\mu}}F(\mu\rho, \sigma) && (50)\\
&\geqslant \frac{1}{2\sqrt{\mu}}\mathrm{Tr}[\mu\rho + \sigma - |\mu\rho - \sigma|] && (51)\\
&= \frac{1}{2\sqrt{\mu}}\mathrm{Tr}[\mu\rho + \sigma - Q(\mu\rho - \sigma) + Q^\perp(\mu\rho - \sigma)] && (52)\\
&= \frac{1}{\sqrt{\mu}}\mathrm{Tr}[Q\sigma + \mu Q^\perp\rho] && (53)\\
&\geqslant \sqrt{\mu}\mathrm{Tr}[Q^\perp\rho] && (54)\\
&= \sqrt{\mu}\epsilon && (55)\\
&\geqslant \epsilon\sqrt{\mu - \mathrm{Tr}[X]/(1-\epsilon)} && (56)\\
&= \epsilon 2^{-\frac{1}{2}D_H^{1-\epsilon}(\rho||\sigma)}. && (57)
\end{aligned}$$

We have used that $||\sqrt{A}\sqrt{B}||_1 \geqslant \mathrm{Tr}[A + B - |A - B|]/2$ for positive semidefinite A, B (a variation of the trace distance bound on the fidelity; see Lemma A.2.6 of [8]).

Now we prove the upper bound. Let Q be primal-optimal for $2^{-D_H^{1-\epsilon}(\rho||\sigma)}$, define $\tilde{\rho} := Q^{\frac{1}{2}}\rho Q^{\frac{1}{2}}$, and let ρ_{AB} be an arbitrary purification of ρ_A. Conjugating both sides of $\rho_{AB} \leqslant \mathbb{I}$ by $Q^{\frac{1}{2}}$, we obtain $\tilde{\rho}_{AB} \leqslant Q_A \otimes \mathbb{I}_B$.

The fidelity between ρ and $\tilde{\rho}$ can be written also in terms of an SDP for $F(\rho_A, \sigma_A)^2$ (with ρ_{AB} an arbitrary purification of ρ_A):

PRIMAL		DUAL	
maximize	$\mathrm{Tr}[\rho_{AB}X_{AB}]$	minimize	$\mathrm{Tr}[Z\sigma]$
subj. to	$\mathrm{Tr}_B[X_{AB}] = \sigma_A$	subj. to	$\rho_{AB} \leqslant Z_A \otimes \mathbb{I}_B$
			$Z \geqslant 0$

We see that Q is a feasible Z_A in the SDP for $F(\tilde{\rho}, \sigma)^2$. Hence,

$$\begin{aligned} 2^{-D_{\min}(\tilde{\rho}||\sigma)} &= F(\tilde{\rho}, \sigma)^2 && (58)\\ &\leqslant \mathrm{Tr}[Q\sigma] && (59)\\ &= (1-\epsilon)2^{-D_H^{(1-\epsilon)}(\rho||\sigma)}, && (60) \end{aligned}$$

and so $D_{\min}(\tilde{\rho}||\sigma) \geqslant D_H^{(1-\epsilon)}(\rho||\sigma) + \log\frac{1}{1-\epsilon}$.

From complementary slackness we get that $\mathrm{Tr}[Q\rho] = 1-\epsilon$. Using Lemma 2 we obtain $P(\tilde{\rho}, \rho) \leqslant \sqrt{1 - \mathrm{Tr}[Q\rho]^2} \leqslant \sqrt{2\epsilon}$, and the first part of the proposition follows.

Rewriting this for $H_{\max}$ and $H_H^{(1-\epsilon)}$ yields:

$$\begin{aligned} H_{\max}(A|B)_\rho &\geqslant H_{\max}(A|B)_{\rho|\rho} && (61)\\ &= -D_{\min}(\rho_{AB}||\mathbb{I}_A \otimes \rho_B) && (62)\\ &\geqslant -D_H^{1-\epsilon}(\rho_{AB}||\mathbb{I}_A \otimes \rho_B) - \log\frac{1}{\epsilon^2} && (63)\\ &= H_H^{(1-\epsilon)}(A|B)_\rho - \log\frac{1}{\epsilon^2} && (64) \end{aligned}$$

□

5 Decomposition of Hypothesis Tests & Entropic Chain Rules

In this section we prove a bound on hypothesis testing between ρ and σ in terms of hypothesis tests between ρ and some other state ξ and ξ and σ. This bound yields a chain rule for the hypothesis testing entropy.

We first require the following Lemma:

Lemma 1. *Let $\tilde{\rho}, \rho \in \mathcal{S}(\mathcal{H})$ be such that $||\tilde{\rho} - \rho||_1 \leqslant \delta$ for some $\delta \geqslant 0$. Then, for any $\sigma \in \mathcal{P}(\mathcal{H})$,*

$$D_H^{\epsilon+\delta}(\rho||\sigma) + \log\frac{\epsilon}{\epsilon+\delta} \leqslant D_H^{\epsilon}(\tilde{\rho}||\sigma). \quad (65)$$

Proof. Let Q be primal-optimal for $D_H^{\epsilon+\delta}(\rho||\sigma)$. It follows that

$$\begin{aligned}\mathrm{Tr}[Q\tilde{\rho}] &= \mathrm{Tr}[Q(\tilde{\rho}-\rho)] + \mathrm{Tr}[Q\rho] && (66)\\ &\geqslant -\delta + \epsilon + \delta && (67)\\ &= \epsilon. && (68)\end{aligned}$$

Hence, Q is primal-feasible for $D_H^{\epsilon}(\tilde{\rho}||\sigma)$, yielding a bound of

$$\begin{aligned}2^{-D_H^{\epsilon}(\tilde{\rho}||\sigma)} &\leqslant \frac{1}{\epsilon}\mathrm{Tr}[Q\sigma] && (69)\\ &= \frac{\epsilon+\delta}{\epsilon}2^{-D_H^{\epsilon+\delta}(\rho||\sigma)}, && (70)\end{aligned}$$

which proves the lemma. □

Now we can state the main result, which deals with the relative entropy of arbitrary states to those that are invariant under a group action. For a group G and unitary representation U_g, let $\mathcal{E}_G(\rho) = \frac{1}{|G|}\sum_{g\in G} U_g\rho U_g^\dagger$, which is a quantum operation. (For simplicity of presentation we assume the group is finite, but the argument applies to continuous groups as well.)

Proposition 9. *For any $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ and group G such that $\sigma = \mathcal{E}_G(\sigma)$, let $\xi = \mathcal{E}_G(\rho)$. Then, for $\epsilon, \epsilon' > 0$,*

$$D_H^{\epsilon+\sqrt{8\epsilon'}}(\rho||\sigma) \leqslant D_H^{\epsilon}(\rho||\xi) + D_H^{\epsilon'}(\xi||\sigma) + \log\frac{\epsilon+\sqrt{8\epsilon'}}{\epsilon}. \quad (71)$$

Proof. Let μ_1 and X_1 be optimal in the dual program of $D_H^{\epsilon}(\rho||\xi)$ and, similarly, μ_2 and X_2 be optimal in $D_H^{\epsilon'}(\xi||\sigma)$. Thus, $\mu_1\rho \leqslant \xi + X_1$ and $\mu_2\xi \leqslant \sigma + X_2$. Observe that X_2 can be chosen G-invariant without loss of generality, since $\mu_2\xi \leqslant \sigma + \mathcal{E}_G(X_2)$ and $\mathrm{Tr}[X_2] = \mathrm{Tr}[\mathcal{E}_G(X_2)]$.

Chaining the inequalities gives

$$\mu_1\mu_2\rho \leqslant \sigma + X_2 + \mu_2 X_1. \quad (72)$$

Next, define $T = \sigma^{\frac{1}{2}}(\sigma + X_2)^{-\frac{1}{2}}$ and conjugate both sides of the above by T. This gives

$$\mu_1\mu_2 T\rho T^\dagger \leqslant \sigma + \mu_2 T X_1 T^\dagger. \quad (73)$$

Thus, the pair $\mu_1\mu_2$, $\mu_2 T X_1 T^\dagger$ is feasible for $D_H^{\epsilon}(T\rho T^\dagger||\sigma)$. Since T is a contraction ($TT^\dagger \leqslant \mathbb{I}$), we can proceed as follows:

$$\begin{aligned}2^{-D_H^{\epsilon}(T\rho T^\dagger||\sigma)} &\geqslant \mu_1\mu_2 - \frac{\mu_2\mathrm{Tr}[TX_1T^\dagger]}{\epsilon} && (74)\\ &\geqslant \mu_1\mu_2 - \frac{\mu_2\mathrm{Tr}X_1}{\epsilon} && (75)\\ &= \mu_2 2^{-D_H^{\epsilon}(\rho||\xi)} && (76)\\ &\geqslant 2^{-D_H^{\epsilon'}(\xi||\sigma)}2^{-D_H^{\epsilon}(\rho||\xi)}. && (77)\end{aligned}$$

Now we show that $P(\rho, T\rho T^\dagger) \leqslant \sqrt{2\epsilon'}$, in order to invoke Lemma 1. Let the isometry $V : \mathcal{H}_A \to \mathcal{H}_A \otimes \mathcal{H}_R$ be a Stinespring dilation of $\mathcal{E}_G$, so that $\overline{\xi}_{AR} = V_{A\to AR}\rho_A V_{A\to AR}^\dagger =$

$\frac{1}{|G|}\sum_{g,g'\in G} U_g \rho U_{g'}^\dagger \otimes |g\rangle\langle g'|$. The state $\overline{\xi}_{AR}$ is an extension of ξ_A since $\xi_A = \mathrm{Tr}_R[\overline{\xi}_{AR}]$. Clearly, $T_A \overline{\xi}_{AR} T_A^\dagger$ is an extension of $T\xi T^\dagger$. We now apply Lemma 3 to the inequality $\xi \leqslant \sigma/\mu_2 + X_2/\mu_2$ to find

$$P(\bar{\xi}_{AR}, T_A \bar{\xi}_{AR} T_A^\dagger) \leqslant \sqrt{\frac{\mathrm{Tr}[X_2]}{\mu_2}\left(2 - \frac{\mathrm{Tr}[X_2]}{\mu_2}\right)} \tag{78}$$

$$= \sqrt{2\epsilon'}. \tag{79}$$

This entails that

$$P(\rho, T\rho T^\dagger) = P(V\rho_A V^\dagger, VT\rho T^\dagger V^\dagger) \tag{80}$$

$$= P(V\rho_A V^\dagger, TV\rho V^\dagger T^\dagger) \tag{81}$$

$$= P(\bar{\xi}_{AR}, T_A \bar{\xi}_{AR} T_A^\dagger) \tag{82}$$

$$\leqslant \sqrt{2\epsilon'}, \tag{83}$$

where we have used the fact that T_A commutes with V_{AR}. This then implies that $||\rho - T\rho T^\dagger||_1 \leqslant \sqrt{8\epsilon'}$. Lemma 1 and (77) then yields the proposition:

$$D_H^{\epsilon+\sqrt{8\epsilon'}}(\rho||\sigma) + \log\frac{\epsilon}{\epsilon+\sqrt{8\epsilon'}} \leqslant D_H^\epsilon(WT\rho T^\dagger W^\dagger||\sigma) \tag{84}$$

$$\leqslant D_H^\epsilon(\rho||\xi) + D_H^{\epsilon'}(\xi||\sigma). \tag{85}$$

□

Corollary 1 (Chain rule for H_H^ϵ)**.** *Let $\rho_{ABC} \in \mathcal{S}(\mathcal{H})$ be an arbitrary normalized state, and $\epsilon, \epsilon' > 0$. Then,*

$$H_H^{\epsilon+\sqrt{8\epsilon'}}(AB|C)_\rho \geqslant H^\epsilon(A|BC)_\rho + H^{\epsilon'}(B|C)_\rho - \log\frac{\epsilon+\sqrt{8\epsilon'}}{\epsilon}. \tag{86}$$

Proof. Let G be the Weyl-Heisenberg group representation (as in the proof of Prop 4) acting on A, for which $\mathcal{E}_G(\rho_{ABC}) = \pi_A \otimes \rho_{BC}$, where $\pi_A = \mathbb{I}/\dim(\mathcal{H}_A)$. Applied to the hypothesis test between ρ_{ABC} and $\pi_{AB}\otimes\rho_C$, we find

$$D_H^{\epsilon+\sqrt{8\epsilon'}}(\rho_{ABC}||\pi_{AB}\otimes\rho_C)$$

$$\leqslant D_H^\epsilon(\rho_{ABC}||\pi_A\otimes\rho_{BC}) + D_H^{\epsilon'}(\pi_A\otimes\rho_{BC}||\pi_{AB}\otimes\rho_C) + \log\frac{\epsilon+\sqrt{8\epsilon'}}{\epsilon} \tag{87}$$

$$\leqslant D_H^\epsilon(\rho_{ABC}||\pi_A\otimes\rho_{BC}) + D_H^{\epsilon'}(\rho_{BC}||\pi_B\otimes\rho_C) + \log\frac{\epsilon+\sqrt{8\epsilon'}}{\epsilon}. \tag{88}$$

As $H_H^\epsilon(A|B)_\sigma = \log d_A - D_H^\epsilon(\sigma_{AB}||\pi_A\otimes\sigma_B)$, this is equivalent to the desired result. □

Acknowledgements

We acknowledge discussions with Marco Tomamichel. Research leading to these results was supported by the Swiss National Science Foundation (through the National Centre of Competence in Research 'Quantum Science and Technology' and grant No. 200020-135048) and by the European Research Council (grant 258932).

A Useful Lemmas

Lemma 2 (Lemma 7, Berta *et al.* [55])**.** *For any $\rho \in \mathcal{S}_{\leqslant}(\mathcal{H})$, and for any nonnegative operator $\Pi \leqslant \mathbb{I}$,*

$$P(\rho, \Pi\rho\Pi) \leqslant \frac{1}{\sqrt{\mathrm{Tr}\rho}}\sqrt{(\mathrm{Tr}\rho)^2 - (\mathrm{Tr}(\Pi^2\rho))^2} \tag{89}$$

Proof. Since $||\sqrt{\rho}\sqrt{\Pi\rho\Pi}||_1 = \mathrm{Tr}\sqrt{(\sqrt{\rho}\Pi\sqrt{\rho})(\sqrt{\rho}\Pi\sqrt{\rho})} = \mathrm{Tr}(\Pi\rho)$, we can write the generalized fidelity as

$$\bar{F}(\rho, \Pi\rho\Pi) = \mathrm{Tr}(\Pi\rho) + \sqrt{(1 - \mathrm{Tr}\rho)(1 - \mathrm{Tr}(\Pi^2\rho))}. \tag{90}$$

For simplicity, introduce the following abbreviations: $r = \mathrm{Tr}\rho$, $s = \mathrm{Tr}(\Pi\rho)$ and $t = \mathrm{Tr}(\Pi^2\rho)$. As $\rho \leqslant 1$ and $\Pi \leqslant 1$ trivially $0 \leqslant t \leqslant s \leqslant r \leqslant 1$. In terms of these variables, we now have that

$$1 - \bar{F}(\rho, \Pi\rho\Pi)^2 = r + t - rt - s^2 - 2s\sqrt{(1-r)(1-t)}. \tag{91}$$

Since $P(\rho, \Pi\rho\Pi) = \sqrt{1 - \bar{F}(\rho, \Pi\rho\Pi)^2}$, it is sufficient to show that $r(1-\bar{F}(\rho, \Pi\rho\Pi)^2) - r^2 + t^2 \leqslant 0$. This we can establish:

$$\begin{aligned}
r(1 - \bar{F}(\rho, \Pi\rho\Pi)^2) - r^2 + t^2 &= r(r + t - rt - s^2 - 2s\sqrt{(1-r)(1-t)}) - r^2 + t^2 && (92)\\
&\leqslant r(r + t - rt - s^2 - 2s(1-r)) - r^2 + t^2 && (93)\\
&= rt - r^2t + t^2 - 2rs + 2r^2s - rs^2 && (94)\\
&\leqslant rt - r^2t + t^2 - 2rs + 2r^2s - rt^2 && (95)\\
&= (1-r)(t^2 + rt - 2rs) && (96)\\
&\leqslant (1-r)(s^2 + rs - 2rs) && (97)\\
&= (1-r)s(s-r) && (98)\\
&\leqslant 0 && (99)
\end{aligned}$$

and the lemma follows. □

Lemma 3 (Lemma 15, Tomamichel *et al.* [56]; Lemma 6.1 [57])**.** *Let $\rho \in \mathcal{S}(\mathcal{H})$, $\sigma \in \mathcal{P}(\mathcal{H})$, $\rho \leqslant \sigma + \Delta$, and $G := \sigma^{\frac{1}{2}}(\sigma + \Delta)^{-\frac{1}{2}}$, where the inverse is taken on the support of σ. Furthermore, let $|\psi\rangle \in \mathcal{S}(\mathcal{H} \otimes \mathcal{H})$ be a purification of ρ. Then,*

$$P(\psi, (G \otimes \mathbb{I})\psi(G^\dagger \otimes \mathbb{I})) \leqslant \sqrt{\mathrm{Tr}\Delta(2 - \mathrm{Tr}\Delta)}. \tag{100}$$

Proof. Let $|\psi\rangle \in \mathcal{S}(\mathcal{H} \otimes \mathcal{H})$ be a purification of ρ. Then, $(G \otimes \mathbb{I})|\psi\rangle$ is a purification of $G\rho G^\dagger$, and with the help of Uhlmann's theorem we can bound the fidelity:

$$\begin{aligned}
F(\psi, (G \otimes \mathbb{I})\rho(G^\dagger \otimes \mathbb{I})) &= |\langle\psi| G \otimes \mathbb{I} |\psi\rangle| && (101)\\
&\geqslant \mathcal{R}\{\mathrm{Tr}(G\rho)\} = \mathrm{Tr}(\bar{G}\rho), && (102)
\end{aligned}$$

with $\bar{G} := \frac{1}{2}(G + G^\dagger)$. Since G is a contraction[6], $||G|| \leqslant 1$. Also, $||\bar{G}|| \leqslant 1$ by the triangle

[6] to see this, conjugate both sides of $\sigma \leqslant \sigma + \Delta$ by $(\sigma + \Delta)^{-1/2}$ to get $G^\dagger G \leqslant 1$.

inequality and thus $\mathrm{Tr}(\bar{G}\rho) \leqslant 1$. Furthermore,

$$\begin{aligned} 1 - \mathrm{Tr}(\bar{G}\rho) &= \mathrm{Tr}((\mathbb{I} - \bar{G})\rho) && (103)\\ &\leqslant \mathrm{Tr}(\sigma + \Delta) - \mathrm{Tr}(\bar{G}(\sigma + \Delta)) && (104)\\ &= \mathrm{Tr}(\sigma + \Delta) - \mathrm{Tr}((\sigma + \Delta)^{\frac{1}{2}}(\sigma)^{\frac{1}{2}}) && (105)\\ &\leqslant \mathrm{Tr}(\Delta), && (106) \end{aligned}$$

where we have used $\rho \leqslant \sigma + \Delta$ and $\sqrt{\sigma + \Delta} \geqslant \sqrt{\sigma}$. Then we find

$$\begin{aligned} P(\psi, (G \otimes \mathbb{I})\psi(G^\dagger \otimes \mathbb{I})) &= \sqrt{1 - F(\psi, (G \otimes \mathbb{I})\psi(G^\dagger \otimes \mathbb{I}))^2} && (107)\\ &\leqslant \sqrt{1 - (1 - \mathrm{Tr}(\Delta)^2)} && (108)\\ &= \sqrt{\mathrm{Tr}\Delta(2 - \mathrm{Tr}\Delta)}. && (109) \end{aligned}$$

□

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[2] J. v. Neumann, *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, 1996.

[3] T. Ogawa and H. Nagaoka, "Strong converse and stein's lemma in quantum hypothesis testing," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2428 –2433, 2000.

[4] H. Nagaoka and M. Hayashi, "An information-spectrum approach to classical and quantum hypothesis testing for simple hypotheses," *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 534 –549, 2007.

[5] T. S. Han, *Information-Spectrum Method in Information Theory*. Springer-Verlag, 2002.

[6] G. Bowen and N. Datta, "Beyond i.i.d. in quantum information theory," in *2006 IEEE International Symposium on Information Theory*, pp. 451–455, 2006.

[7] R. Renner and S. Wolf, "Smooth Renyi Entropy and Applications," in *2004 IEEE International Symposium on Information Theory*, pp. 232–232, IEEE, 2004.

[8] R. Renner, *Security of quantum key distribution*. PhD thesis, ETH Zurich, 2005. arXiv:quant-ph/0512258.

[9] N. Datta and R. Renner, "Smooth Entropies and the Quantum Information Spectrum," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2807–2815, 2009.

[10] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review*, vol. 106, no. 4, p. 620, 1957.

[11] J. Aczél, B. Forte, and C. T. Ng, "Why the shannon and hartley entropies are 'Natural'," *Advances in Applied Probability*, vol. 6, no. 1, p. 131, 1974.

[12] W. Ochs, "A new axiomatic characterization of the von neumann entropy," *Reports on Mathematical Physics*, vol. 8, no. 1, pp. 109–120, 1975.

[13] E. H. Lieb and J. Yngvason, "A guide to entropy and the second law of thermodynamics," *Notices of the American Mathematical Society*, vol. 45, no. 5, p. 571, 1998.

[14] E. H. Lieb and J. Yngvason, "A fresh look at entropy and the second law of thermodynamics," *Physics Today*, vol. 53, no. 4, pp. 32–37, 2000.

[15] I. Csiszár, "Axiomatic characterizations of information measures," *Entropy*, vol. 10, no. 3, pp. 261–273, 2008.

[16] B. Baumgartner, "Characterizing entropy in statistical physics and in quantum information theory." arXiv:1206.5727, 2012.

[17] R. Landauer, "Irreversibility and heat generation in the computing process," *IBM Journal of Research and Development*, vol. 5, no. 3, p. 183, 1961.

[18] C. Bennett, "Logical reversibility of computation," *IBM Journal of Research and Development*, vol. 17, no. 6, p. 525, 1973.

[19] L. del Rio, J. Åberg, R. Renner, O. C. O. Dahlsten, and V. Vedral, "The thermodynamic meaning of negative entropy," *Nature*, vol. 474, no. 7349, pp. 61–63, 2011.

[20] O. C. O. Dahlsten, R. Renner, E. Rieper, and V. Vedral, "Inadequacy of von Neumann entropy for characterizing extractable work," *New Journal of Physics*, vol. 13, no. 5, p. 053015, 2011.

[21] P. Faist, F. Dupuis, J. Oppenheim, and R. Renner, "A quantitative landauer's principle." arXiv:1211.1037, 2012.

[22] R. Renner and R. König, "Universally composable privacy amplification against quantum adversaries," in *Theory of Cryptography*, vol. 3378 of *Lecture Notes in Computer Science*, pp. 407–425, Springer, 2005.

[23] J. M. Renes and R. Renner, "One-Shot Classical Data Compression With Quantum Side Information and the Distillation of Common Randomness or Secret Keys," *IEEE Transactions on Information Theory*, vol. 58, pp. 1985–1991, 2012.

[24] F. Dupuis, *The decoupling approach to quantum information theory*. PhD thesis, Université de Montréal, 2009. arXiv:1004.1641.

[25] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner, "The Decoupling Theorem." arXiv:1012.6044, 2010.

[26] M. Horodecki, J. Oppenheim, and A. Winter, "Partial quantum information.," *Nature*, vol. 436, no. 7051, pp. 673–6, 2005.

[27] M. Horodecki, J. Oppenheim, and A. Winter, "Quantum State Merging and Negative Information," *Communications in Mathematical Physics*, vol. 269, no. 1, pp. 107–136, 2006.

[28] M. Berta, "Single-shot Quantum State Merging," Master's thesis, ETH Zurich, 2008. arXiv:0912.4495.

[29] R. Renner, S. Wolf, and J. Wullschleger, "The Single-Serving Channel Capacity," in *2006 IEEE International Symposium on Information Theory*, pp. 1424–1427, IEEE, 2006.

[30] J. M. Renes and R. Renner, "Noisy Channel Coding via Privacy Amplification and Information Reconciliation," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7377–7385, 2011.

[31] M. Berta, M. Christandl, and R. Renner, "The Quantum Reverse Shannon Theorem Based on One-Shot Information Theory," *Communications in Mathematical Physics*, vol. 306, no. 3, pp. 579–615, 2011.

[32] I. B. Damgård, S. Fehr, R. Renner, L. Salvail, and C. Schaffner, "A tight high-order entropic quantum uncertainty relation with applications," in *Advances in Cryptology - CRYPTO 2007*, vol. 4622, pp. 360–378, Springer, 2007.

[33] V. Scarani and R. Renner, "Quantum Cryptography with Finite Resources: Unconditional Security Bound for Discrete-Variable Protocols with One-Way Postprocessing," *Physical Review Letters*, vol. 100, no. 20, pp. 1–4, 2008.

[34] J. Wullschleger, "Oblivious-Transfer amplification," in *Advances in Cryptology—EUROCRYPT '07*, vol. 4515 of *Lecture Notes in Computer Science*, pp. 555–572, Springer, 2007.

[35] I. B. Damgård, S. Fehr, R. Renner, L. Salvail, and C. Schaffner, "A Tight High-Order Entropic Quantum Uncertainty Relation With Applications," in *Advances in Cryptology—CRYPTO '07*, vol. 4622 of *Lecture Notes in Computer Science*, pp. 360–378, Springer, 2007.

[36] F. Buscemi and N. Datta, "The quantum capacity of channels with arbitrarily correlated noise," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1447–1460, 2010.

[37] F. Brandão and N. Datta, "One-shot rates for entanglement manipulation under non-entangling maps," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1754–1760, 2011.

[38] L. Wang and R. Renner, "One-shot classical-quantum capacity and hypothesis testing," *Physical Review Letters*, vol. 108, no. 20, p. 200501, 2012.

[39] M. Tomamichel and M. Hayashi, "A hierarchy of information quantities for finite block length analysis of quantum tasks." arXiv:1208.1478, 2012.

[40] S. Kullback and R. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[41] A. Wehrl, "General properties of entropy," *Reviews of Modern Physics*, vol. 50, no. 2, pp. 221–260, 1978.

[42] C. W. Helstrom, "Quantum Detection and Estimation Theory," *Journal of Statistical Physics*, vol. 1, no. 2, pp. 231–252, 1969.

[43] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[44] A. Gilchrist, N. K. Langford, and M. A. Nielsen, "Distance measures to compare real and ideal quantum processes," *Physical Review A*, vol. 71, no. 6, p. 062310, 2005.

[45] A. E. Rastegin, "Sine distance for quantum states." arXiv:quant-ph/0602112, 2006.

[46] M. Tomamichel, R. Colbeck, and R. Renner, "Duality between smooth min- and max-entropies," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4674–4681, 2010.

[47] C. Fuchs and J. van de Graaf, "Cryptographic distinguishability measures for quantum-mechanical states," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1216–1227, 1999.

[48] J. Watrous, "Semidefinite programs for completely bounded norms," *Theory of Computing*, vol. 5, pp. 217–238, 2009.

[49] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[50] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric algorithms and combinatorial optimization*. Springer-Verlag, 1993.

[51] A. Rényi, "On measures of entropy and information," in *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 547–561, 1961.

[52] R. König, R. Renner, and C. Schaffner, "The Operational Meaning of Min- and Max-Entropy," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4337–4347, 2009.

[53] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner, "Leftover Hashing Against Quantum Side Information," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5524–5535, 2011.

[54] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability," *Communications in Mathematical Physics*, vol. 143, no. 1, pp. 99–114, 1991.

[55] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, "The uncertainty principle in the presence of quantum memory," *Nature Physics*, vol. 6, pp. 659–662, 2010.

[56] M. Tomamichel, R. Colbeck, and R. Renner, "A fully quantum asymptotic equipartition property," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5840–5847, 2009.

[57] M. Tomamichel, *A Framework for Non-Asymptotic Quantum Information Theory*. PhD, ETH Zurich, 2012. arXiv:1203.2142.